

O/P/E

## CRF Problem Report

The Scientific and Technical Information Center (STIC) experienced a problem when processing the following computer readable form (CRF):

Application Serial Number: 09/813775B  
Filing Date: 3/20/2001  
Date Processed by STIC: 1/26/2002

STIC Contact: Mark Spencer, 703-308-4212

### Nature of Problem:

The CRF (was):

- ☒ (circle one) Damaged or Unreadable (for Unreadable, see attached)  
☐ Blank (no files on CRF) (see attached)  
☐ Empty file (filename present, but no bytes in file) (see attached)  
☐ Virus-infected. Virus name: \_\_\_\_\_ The STIC will not process the CRF.  
☐ Not saved in ASCII text  
☐ Sequence Listing was embedded in the file. According to Sequence Rules, submitted file should **only** be the Sequence Listing.  
☐ Did not contain a Sequence Listing. (see attached sample)  
☐ Other: \_\_\_\_\_

PLEASE USE THE CHECKER VERSION 3.1 PROGRAM TO REDUCE ERRORS.  
SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:  
U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7<sup>th</sup> Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202  
Or  
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service , or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 01/29/2002

Dear Examiner,

The attached search was run with the most recently released version of Compugen's search software, GenCore 5. With this update, several changes have occurred in the results of FrameSearches (protein query sequence vs nucleic acid databases or nucleic acid query sequence vs protein databases).

In reference to FrameSearches:

- The output format has been improved so that it more closely resembles the format for standard search output.
- Calculation of Percent Similarity has been changed for FrameSearches. The new method of calculation is more similar to the method used in NCBI's BLAST algorithm. The same results are found in the same order using GenCore 5 and the previous version of GenCore, but Percent Similarities are lower in GenCore 5 results.
- The formula for % similarity calculation is:

$$100 * \frac{\text{matches} + \text{conservative substitutions}}{\text{alignment\_length}}$$

where "matches" is the number of identical matches and "conservative substitutions" is the number of non-identical positive matches.

- GenCore 4.5 considers the match Thr vs GCT (Ala) to be a similarity since BLOSUM62 gives score of 0 to this match. It is marked by ':::' in the alignment:

```
Qy      46 AspSerThrAspAla.Met..Gly 52
          |||||:::  ||| :::  |||
Db     605 GATTCCGCTGCTGCTAATTTGGC 628
```

GenCore 5 requires a positive score to consider a non-identical match a similarity, therefore the same 'match' is not emphasized in the new alignment:

```
Qy      46 AspSerThrAspAla.Met..Gly 52
          |||||  ||| :::  |||
Db     605 GATTCCGCTGCTGCTAATTTGGC 628
```

If you have any questions, please feel free to contact one of the searchers in Biotech/Chem Library.